

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ



Μαθηματική Στατιστική -
Εργασίες

Σάμης Τρέβεζας

Σχολή Θετικών Επιστημών
Τμήμα Μαθηματικών
2020

Εργασία 10

Η φοιτητική κοινότητα είναι αναστατωμένη. Ένας άνθρωπος βρέθηκε νεκρός μέσα στο χώρο της Πανεπιστημιούπολης. Το περιστατικό αυτό συνέβει την ώρα των μαθημάτων λίγο πριν την απαγόρευση των δια ζώσης μαθημάτων. Τα μαθήματα σταματάνε. Η αστυνομία έχει εισέρθει μέσα στο χώρο της Πανεπιστημιούπολης προκαλώντας πανικό. Μετά από ένα ανώνυμο τηλεφώνημα η αστυνομία έχει πληροφορηθεί ότι ο Μπάμπης του Μαθηματικού βρίσκεται πίσω από την άγρια δολοφονία του αγνώστου. Το μυστήριο μεγαλώνει καθώς ο Μπάμπης είναι άφαντος. Οι φίλοι του βρίσκονται σε κατάσταση σοκ. Από τη μία ο Μπάμπης ήταν μαζί τους σήμερα την ώρα των μαθημάτων και από την άλλη κάποια στιγμή τον έχασαν χωρίς να ξέρουν που έχει πάει. Βγαίνει ένταλμα σύλληψης και όλοι είναι συγκλονισμένοι. Η ομάδα εξιχνίασης εγκλημάτων της αστυνομίας έχει αποκλείσει το χώρο του συμβάντος. Μετά από αρκετή ώρα δουλειάς έγινε συλλογή πολλών δειγμάτων στα οποία θα γίνει ανάλυση DNA. Οι μέρες πέρασαν και ο Μπάμπης εξακολουθεί να είναι άφαντος. Τα αποτελέσματα DNA βγήκαν και σε ένα δείγμα του θύματος έχει βρεθεί μικτό DNA. Οι ειδικοί μπορούν να ξεχωρίσουν ότι στο δείγμα αυτό υπάρχει το DNA του θύματος και ενός αγνώστου, το οποίο ταυτοποιήθηκε σε 5 τόπους (loci). Θα είναι το δείγμα αυτό ικανοποιητικό για την εξιχνίαση του εγκλήματος ? Οι φίλοι του Μπάμπη δεν μπορούν να πιστέψουν ότι ο φίλος τους έχει κατηγορηθεί και από την άλλη είναι εξαφανισμένος. Θέλουν να μπορέσουν να τον βοηθήσουν για την αθώωσή του έτσι ώστε να ενταθούν οι έρευνες για τον εντοπισμό του. Δεν μπορούν να πιστέψουν ότι όλοι θεωρούν ότι ο Μπάμπης κρύβεται για να μη συλληφθεί. Οι φίλοι του είναι σίγουροι για την αθωότητά του. Παίρνοντας ένα μάθημα Στατιστικής έχουν στοιχειώδεις γνώσεις και γνωρίζουν ότι μπορούν να διεξάγουν έναν έλεγχο με το λόγο πιθανοφάνειας για να μπορέσουν να δώσουν στοιχεία υπέρ της υπεράσπισής του. Το πρόβλημα είναι ότι ο Μπάμπης είναι εξαφανισμένος και αυτό κάνει πιο δύσκολη τη δουλειά τους. Τελικά τα κατάφεραν. Η ομάδα εγκληματολογίας τους έδωσε τα στοιχεία που χρειαζόνταν. Χωρίς να υπάρχει διαθέσιμο δείγμα DNA του Μπάμπη, οι φίλοι του σκέφτονται ότι μπορούν να χρησιμοποιήσουν αυτό των γονιών του για τον έλεγχο. Μπορούν επίσης να έχουν πρόσβαση σε μία βάση δεδομένων από 10.000 δείγματα DNA για να καθορίσουν τις συχνότητες εμφάνισης των χαρακτηριστικών που εξετάζουν.

Σε αυτήν την εργασία ζητείται να αναλάβετε εσείς το ρόλο των φίλων του Μπάμπη στη

διεξαγωγή του ελέγχου λόγου πιθανοφάνειας. Στην εγκληματολογία, σε αρκετές περιοχές τα στοιχεία αυτά παρουσιάζονται στο δικαστήριο ως βοηθητικά για τη διαλεύκανση μιας υπόθεσης (επισυνάπτεται οδηγός από το Ευρωπαϊκό δίκτυο Ινστιτούτων Εγκληματολογίας; ENFSI). Αν

H_0 : “το DNA που ταυτοποιήθηκε θα μπορούσε να είναι οποιοδήποτε άλλου”

vs

H_1 : “το DNA που ταυτοποιήθηκε είναι του Μπάμπη”

τότε ζητείται να πραγματοποιηθεί ένας κατάλληλος έλεγχος λόγου πιθανοφάνειας για να χρησιμοποιηθεί ως βοηθητικό στοιχείο [τυπικά η μηδενική και η εναλλακτική μπορεί να είναι λίγο πιο πολύπλοκες περιλαμβάνοντας μία μίξη DNA με αυτό του θύματος]. Δίνεται η πληροφορία ότι τα 5 loci του DNA που ταυτοποιήθηκε είναι ανεξάρτητα. Κάθε δείγμα DNA θα χαρακτηρίζεται επομένως από ένα συνδυασμό 5 ανεξάρτητων γονιδιακών τύπων, έναν για κάθε loci. Συμβολίζουμε με $G = (G_1, G_2, G_3, G_4, G_5)$ κάθε τέτοιο συνδυασμό. Ο i -γονότυπος G_i διαμορφώνεται από έναν συνδυασμό αλληλόμορφων (allele). Συγκεκριμένα, δίνεται ότι για τα 2 πρώτα, το G_1 και το G_2 γίνονται συνδυασμοί από 2 δυνατά αλληλόμορφα και για τα 3 τελευταία, τα G_3, G_4 και G_5 γίνονται συνδυασμοί από 3 αλληλόμορφα. Γενικά, αν υπάρχουν m δυνατά αλληλόμορφα $A_{i,1}, A_{i,2}, \dots, A_{i,m}$ για τον καθορισμό του γονότυπου G_i , τότε υπάρχουν $m(m+1)/2$ δυνατοί συνδυασμοί ζευγών αλληλόμορφων που σχηματίζονται με την ένωση ενός αλληλόμορφου από τη μητέρα και ενός αλληλόμορφου από τον πατέρα. Η συχνότητα εμφάνισης κάθε αλληλόμορφου καθορίζει και τη συχνότητα εμφάνισης κάθε γονιδιακού τύπου. Συγκεκριμένα, αν γνωρίζουμε ότι $p_{i,j}$ είναι η συχνότητα εμφάνισης του j -αλληλόμορφου του γονιδίου i , τότε εφόσον η επιλογή του ζεύγους των αλληλόμορφων κατά τη γενετική επιλογή γίνεται με τυχαία επιλογή ενός αλληλόμορφου της μητέρας (από τα δύο που έχει) και ενός του πατέρα (από τα δικά του δύο), συμπεραίνουμε ότι οι συχνότητες εμφάνισης κάθε γονιδιακού τύπου $A_{i,j}A_{i,j}$ θα είναι $p_{i,j}^2$ και κάθε γονιδιακού τύπου $A_{i,j}A_{i,k}$ με $j \neq k$ θα είναι $2p_{i,j}p_{i,k}$.

Στο αρχείο που βρίσκεται στην ηλεκτρονική τάξη μπορείται να βρείτε μία βάση δεδομένων που περιλαμβάνει το δείγμα των 10.000 ατόμων, όπως επίσης και τα δείγματα που αντιστοιχούν στον πατέρα και στη μητέρα του Μπάμπη. Βγάλτε λοιπόν τα συμπεράσματά σας για να διαλευκάνετε την υπόθεση. Πιστεύετε ότι είναι αθώος ο Μπάμπης ή έκρυβε μέσα του ένα Μπάμπη Σουγιά ? Συνοδεύστε το αποτέλεσμα σας με κάποιο μέτρο στατιστικής σημαντικότητας. Μπορείτε να συμπεριλάβετε και στοιχεία από τις συστάσεις του ENFSI.